

Synchronizing Objectives for Markov Decision Processes

Laurent Doyen

LSV, ENS Cachan & CNRS, France

doyen@lsv.ens-cachan.fr

Thierry Massart

Mahsa Shirmohammadi

Université Libre de Bruxelles, Brussels, Belgium*

thierry.massart@ulb.ac.be

mahsa.shirmohammadi@ulb.ac.be

Abstract. We introduce synchronizing objectives for Markov decision processes (MDP). Intuitively, a synchronizing objective requires that eventually, at every step there is a state which concentrates almost all the probability mass. In particular, it implies that the probabilistic system behaves in the long run like a deterministic system: eventually, the current state of the MDP can be identified with almost certainty.

We study the problem of deciding the existence of a strategy to enforce a synchronizing objective in MDPs. We show that the problem is decidable for general strategies, as well as for blind strategies where the player cannot observe the current state of the MDP. We also show that pure strategies are sufficient, but memory may be necessary.

1 Introduction

A *Markov decision process (MDP)* is a model for systems that exhibit both probabilistic and nondeterministic behavior. MDPs have been used to model and solve control problems for stochastic systems where the nondeterminism represents the freedom of the controller to choose a control action, while the probabilistic component of the behavior describes the system response to control actions. MDPs have also been adopted as models for concurrent probabilistic systems, probabilistic systems operating in open environments [7], and under-specified probabilistic systems [4].

Traditional objectives for MDP specify a set S of paths, where a path is an infinite sequence of states through the underlying graph of the MDP. The value of interest is the probability that an execution of the MDP under a given strategy belongs to S . For example, a reachability objective specifies all paths that visit a given target state ℓ . A typical qualitative question is to decide whether there exists a strategy such that a given state ℓ is reached with probability 1.

In this paper, we consider a different type of objectives which specify a set of infinite sequences $\bar{X} = X_0, X_1, \dots$ of probability distributions over the states [6]. Intuitively, the distribution X_i in the sequence gives for each state ℓ the probability $X_i(\ell)$ to be in state ℓ at step $i \geq 0$. We introduce *synchronizing objectives* which specify sequences of distributions in which the probability tends to accumulate in a single state. We use the infinity norm as a measure of the highest peak in a probability distribution X_i (i.e., $\|X_i\| = \max_{\ell \in L} X_i(\ell)$) and we require that the limit¹ of this measure in the sequence is 1. Intuitively, this requires that in the long run, the MDP behaves like a deterministic system: from some point on, at every step i there is a state ℓ_i which accumulates almost all the probability. Note that satisfying such an

*This work has been done in the MoVES project (P6/39) which is part of the IAP-Phase VI Interuniversity Attraction Poles Programme funded by the Belgian State, Belgian Science Policy.

¹ Since the limit may not exist in general, we actually consider either \liminf or \limsup .

objective implies that there exists a state ℓ which is reached with probability 1. The converse does not hold because reachability objectives do not require the visits to the target state to occur after the same number of steps in (almost) all executions of the MDP. We consider the problem of deciding if a given MDP is synchronizing for some strategy. We consider the general case where memoryful randomized strategies are allowed, as well as the special case of blind strategies which are not allowed to observe the current state of the MDP.

Defining objectives as a sequence of probability distributions over states rather than a distribution over sequences of states is a change of standpoint in the traditional approach to MDP verification. Up to our knowledge, there are very few works in this setting. We are aware of the work in [6] which studies MDPs as generators of probability distributions with applications in sensor networks and dynamical systems, and shows that the resulting objectives are not expressible in known logics such as PCTL* [1, 4]. In their definition, probability distributions over states are assigned a vector $v \in \{0, 1\}^k$ of truth values for a finite set of predicates $\varphi_1, \dots, \varphi_k$ (which are linear constraints on the probabilities such as $\varphi(X) \equiv X(\ell) + X(\ell') \leq \frac{1}{2}$, for example). This can be viewed as a coloring of the probability distributions using a finite number of colors, and then objectives are languages of infinite words over the finite alphabet of colors. It is shown that reachability of a given color is undecidable for MDPs if arbitrary linear predicates are allowed [6]. A decidability result is obtained if only predicates of the form $\sum_{\ell \in T} X(\ell) > 0$ are allowed. Synchronizing objectives cannot be expressed in the framework of [6] using finite colorings as they require a real-valued measure (namely, the infinite norm) to be assigned to the probability distributions.

In [2], the monadic logic of probabilities is introduced as a predicate logic which can express properties of sequences of probability distributions. But because it allows comparison of probabilities only with constants, it cannot express synchronizing objectives which would require a quantification over probability thresholds, such as $\varphi(\bar{X}) \equiv \forall \varepsilon > 0 \cdot \exists N \cdot \forall i \geq N \cdot \exists \ell \in L : X_i(\ell) \geq 1 - \varepsilon$, where X_i is the probability distribution in position i in the sequence \bar{X} .

Synchronizing objectives generalize the notion of synchronizing words. In a deterministic finite automaton, a word w is synchronizing if reading w from any state of the automaton always leads to the same state. It is sufficient to consider finite words, and it is conjectured that if a synchronizing word exists, then there exists one of length at most $(n - 1)^2$ where n is the number of states of the automaton, known as the Černý's conjecture. Several works have studied this conjecture and related problems (see the survey in [8]). Viewing deterministic automata as a special case of MDP where all transitions have only one successor, a synchronizing word can be seen as a blind strategy to ensure a synchronizing objective. Note that we do not present a generalization of Černý's conjecture since in our case, strategies for MDPs are infinite objects. However, synchronizing objectives provide an extension of the design framework for the many applications of the theory of synchronizing words, such as control of discrete event systems, planning, biocomputing, and robotics [8]. For example, in probabilistic models of DNA transcription, one may ask which molecules to introduce in a cell in order to bring it to a single possible state [3, 8].

We prove that it is decidable to determine if a given MDP is synchronizing for some strategy, either blind or general. We use variants of the subset construction in the underlying graph of MDPs to obtain a decidable characterization of synchronizing strategies. Our results imply that pure strategies are sufficient to satisfy a synchronizing objective, but we provide an example showing that memory may be necessary, both with blind and general strategies.

2 Definitions.

A *probability distribution* over a finite set S is a function $d : S \rightarrow [0, 1]$ such that $\sum_{s \in S} d(s) = 1$. The *support* of d is the set $\text{Supp}(d) = \{s \in S \mid d(s) > 0\}$. $\mathcal{D}(S)$ denotes the set of all probability distributions on S , and $\mathcal{P}(S)$ the power set of S .

Markov decision processes. A *Markov decision process* (MDP) is a tuple $M = \langle L, \mu_0, \Sigma, \delta \rangle$ where L is a finite set of states, $\mu_0 \in \mathcal{D}(L)$ is an initial probability distribution over states, Σ is a finite set of actions, $\delta : L \times \Sigma \rightarrow \mathcal{D}(L)$ is a probabilistic transition function that assigns to each pair of states and actions, a probability distribution over successor states. A *Markov chain* is a special case of MDPs with only one action ($|\Sigma| = 1$). Markov chains are therefore generally viewed as a tuple $M = \langle L, \mu_0, \delta \rangle$ where $\delta : L \rightarrow \mathcal{D}(L)$. For an action $\sigma \in \Sigma$ and a state $\ell \in L$, let $\text{Post}_\sigma(\ell) = \text{Supp}(\delta(\ell, \sigma))$, and for a set $s \subseteq L$, let $\text{Post}_\sigma(s) = \cup_{\ell \in s} \text{Post}_\sigma(\ell)$.

Example Figure 1(a) shows an MDP with four states and alphabet $\Sigma = \{\sigma_1, \sigma_2\}$. The initial probability distribution is $\mu_0(1) = 1$ and $\mu_0(i) = 0$ for $i \in \{2, 3, 4\}$, and the probabilistic transition function δ in state 1 is such that $\delta(1, \sigma_1)(2) = \delta(1, \sigma_1)(3) = 1/2$ and $\delta(1, \sigma_2)(1) = 1$.

We describe the behavior of an MDP as a one-player stochastic game played for infinitely many rounds. In the first round, the game starts in state ℓ with probability $\mu_0(\ell)$. In each round, if the game is in the state ℓ and the player chooses the action $\sigma \in \Sigma$, then the game moves to the successor state ℓ' chosen with probability $\delta(\ell, \sigma)(\ell')$, and the next round starts. We consider two versions of this game. In both versions, the player knows the structure of the MDP. In the first version the player has *perfect information*, he can see the current state of the game; in the second version the player is *blind*, he is not allowed to observe the current state of the game, and only knows the number of rounds that have been played so far.

A *play* of the game is an infinite sequence of interleaved states and actions $\pi = \ell_0 \sigma_0 \ell_1 \cdots$ such that $\ell_{i+1} \in \text{Post}_{\sigma_i}(\ell_i)$ for all $i \geq 0$. The set of all plays over M is denoted by $\text{Plays}(M)$. A finite prefix $h = \ell_0 \sigma_0 \ell_1 \cdots \sigma_{n-1} \ell_n$ of a play π is called a *history*, the last state of h is $\text{Last}(h) = \ell_n$, the i^{th} action and state of the h is $\text{Action}(h, i) = \sigma_i$ and $\text{State}(h, i) = \ell_i$, and its length is $|h| = n$. The set of all histories of plays is denoted by $\text{Hists}(M)$.

Strategies and outcome. In the game, the choice of the action is made by the player according to a strategy. Depending on what the player can observe and record, he can use various classes of strategies. A *randomized strategy* (or simply a strategy) over an MDP M is a function $\alpha : \text{Hists}(M) \rightarrow \mathcal{D}(\Sigma)$. A *pure* (deterministic) strategy is a special case of randomized strategy where for all $h \in \text{Hists}(M)$, there exists an action $\sigma \in \Sigma$ such that $\alpha(h)(\sigma) = 1$. A *memoryless* strategy is a randomized strategy α such that $\alpha(h_1) = \alpha(h_2)$ for all $h_1, h_2 \in \text{Hists}(M)$ with $\text{Last}(h_1) = \text{Last}(h_2)$. In this last case, the player cannot record the history of the play and makes a choice according to the current state only. For convenience, we view pure strategies as functions $\alpha : \text{Hists}(M) \rightarrow \Sigma$, and memoryless strategies as functions $\alpha : L \rightarrow \mathcal{D}(\Sigma)$. Hence, a pure memoryless strategy is a function $\alpha : L \rightarrow \Sigma$.

A strategy α is *blind* if $\alpha(h_1) = \alpha(h_2)$ for all $h_1, h_2 \in \text{Hists}(M)$ such that $|h_1| = |h_2|$. Blind strategies can be viewed as functions $\alpha : \mathbb{N} \rightarrow \mathcal{D}(\Sigma)$ (or, $\alpha : \mathbb{N} \rightarrow \Sigma$ for pure blind strategies) which assign in each round a probability distribution over actions. Sometimes we talk about *perfect-information* strategies to emphasize when we consider strategies that are not necessarily blind.

The *outcome* of the game played on an MDP $M = \langle L, \mu_0, \Sigma, \delta \rangle$ using a strategy α is the infinite sequence $X_0^\alpha X_1^\alpha \dots$ of probability distributions over the set of states L , where $X_0^\alpha = \mu_0$ and for all $n > 0$,

$$X_n^\alpha(\ell) = \sum_{h \in \text{Hists}(M): \text{Last}(h) = \ell, |h| = n} \text{Pr}^\alpha(h)$$

where the probability $\text{Pr}^\alpha(h)$ of a history $h = \ell_0 \sigma_0 \ell_1 \dots \sigma_{n-1} \ell_n$ under strategy α is

$$\text{Pr}^\alpha(h) = \mu_0(\ell_0) \cdot \prod_{j=1}^n \alpha(\ell_0 \sigma_0 \dots \ell_{j-1})(\sigma_{j-1}) \cdot \delta(\ell_{j-1}, \sigma_{j-1})(\ell_j).$$

Synchronizing objectives. The *norm* of a probability distribution X over L is $\|X\| = \max_{\ell \in L} X(\ell)$. We say that the MDP M with strategy α is *strongly synchronizing* if

$$\liminf_{n \rightarrow \infty} \|X_n^\alpha\| = 1, \quad (1)$$

and that it is *weakly synchronizing* if

$$\limsup_{n \rightarrow \infty} \|X_n^\alpha\| = 1. \quad (2)$$

Intuitively, an MDP is synchronizing if the probability mass tends to concentrate in a single state, either at every step from some point on (for strongly synchronizing), or at infinitely many steps (for weakly synchronizing). Note that equivalently, M with strategy α is strongly synchronizing if the limit $\lim_{n \rightarrow \infty} \|X_n^\alpha\|$ exists and equals 1. In this paper, we are interested in the problem of deciding if a given MDP is synchronizing for some strategy. We consider the problem for both perfect-information and blind strategies.

Recurrent and transient states. A state $\ell' \in L$ is *accessible* from a state $\ell \in L$ (denoted $\ell \rightarrow \ell'$), if there is a history $h = \ell_0 \sigma_0 \ell_1 \dots \sigma_{n-1} \ell_n$ with $\ell_0 = \ell$ and $\ell_n = \ell'$. If both $\ell \rightarrow \ell'$ and $\ell' \rightarrow \ell$ hold, then we say that ℓ and ℓ' are *strongly connected* (denoted $\ell \leftrightarrow \ell'$). This induces an equivalence relation called *accessibility relation*. An MDP is *strongly connected*, if all pairs of states $\ell, \ell' \in L$ are strongly connected. A state accessible from a state of $\text{Supp}(\mu_0)$ is simply called *accessible state*.

For a Markov chain M , the state ℓ is *recurrent* if all accessible states from ℓ can access ℓ (i.e., ℓ and ℓ' are strongly connected for all ℓ' such that $\ell \rightarrow \ell'$), and the state ℓ is *transient* if there exists some state ℓ' such that ℓ' is accessible from ℓ , but ℓ is not accessible from ℓ' . The next proposition follows from standard results [5].

Proposition 1 *Given a Markov chain M , let X_0, X_1, \dots be the sequence of probability distributions of M . Then $\limsup_{n \rightarrow \infty} X^n(\ell) = 0$ for all transient states $\ell \in L$, and $\limsup_{n \rightarrow \infty} X^n(\ell) > 0$ for all recurrent states $\ell \in L$.*

Subset constructions. We define two important constructions based on the subset construction idea. Subset construction is a standard technique to compute, from a nondeterministic finite automaton N , an equivalent deterministic automaton D (for language equivalence), where one state of D corresponds to the set of possible states (called a cell) in which N can be. We define two kinds of subset constructions on MDPs, the *perfect-information subset construction*, and the *blind subset construction*. As usual, each state of the subset constructions is a subset of states of the MDP (i.e., a cell). In our case, the main difference lies in the alphabet. In the perfect-information subset construction, the selection of the next action depends on the current state (each state of a cell can independently choose an action), while in the blind subset constructions the next action is independent of the state (all states of a cell have to choose the same action). Thus, an action in the perfect-information subset construction is a function $\hat{\sigma} : L \rightarrow \Sigma$ which assigns to each state $\ell \in L$ its choice among the actions in Σ .

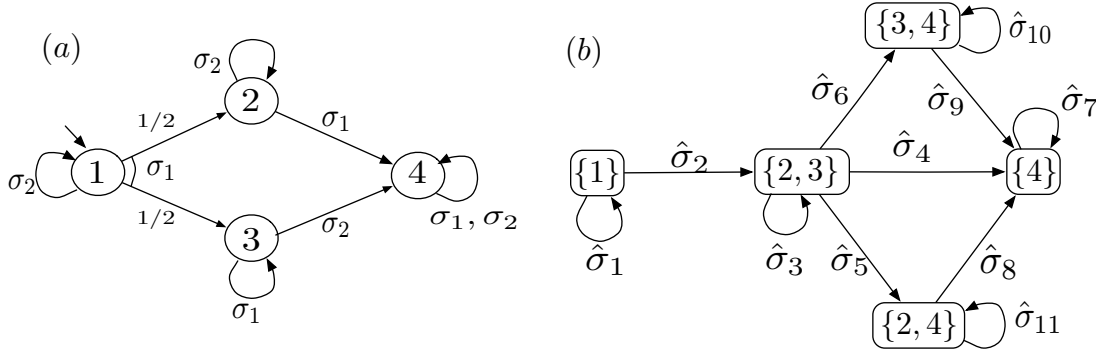


Figure 1: (a) shows an MDP, and (b) shows the accessible states of its perfect information subset construction.

Definition 1 (Perfect-information subset construction of an MDP) For an MDP $M = \langle L, \mu_0, \Sigma, \delta \rangle$, the perfect-information subset construction is an automaton $M^P = \langle \mathcal{L}, L_I, \hat{\Sigma}, \delta^P \rangle$ where $\mathcal{L} = \mathcal{P}(L) \setminus \{\emptyset\}$, $L_I = \text{Supp}(\mu_0)$, $\hat{\Sigma} = \{\hat{\sigma} \mid \hat{\sigma} : L \rightarrow \Sigma\}$ is the alphabet, and $\delta^P : \mathcal{L} \times \hat{\Sigma} \rightarrow \mathcal{L}$ where for all $s_1, s_2 \in \mathcal{L}$ and $\hat{\sigma} \in \hat{\Sigma}$, we have $\delta^P(s_1, \hat{\sigma}) = s_2$ where $s_2 = \bigcup_{\ell \in s_1} \text{Post}_{\hat{\sigma}(\ell)}(\ell)$.

Example Figure 1(b) shows the perfect information subset construction M^P of the MDP drawn in Figure 1(a) (presented in the first example). Let us present $\hat{\Sigma}$ in the table below. Each row labelled by a function $\hat{\sigma}_i$ ($i \in \{1, \dots, 11\}$), each column labelled by a state ℓ ; and each entry shows the value of $\hat{\sigma}_i(\ell)$.

	1	2	3	4
$\hat{\sigma}_1$	σ_2	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_2$	σ_1	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_3$	$\{\sigma_1, \sigma_2\}$	σ_2	σ_1	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_4$	$\{\sigma_1, \sigma_2\}$	σ_1	σ_2	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_5$	$\{\sigma_1, \sigma_2\}$	σ_2	σ_2	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_6$	$\{\sigma_1, \sigma_2\}$	σ_1	σ_1	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_7$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_8$	$\{\sigma_1, \sigma_2\}$	σ_1	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_9$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	σ_2	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_{10}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	σ_1	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_{11}$	$\{\sigma_1, \sigma_2\}$	σ_2	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$

Note that, the function $\hat{\sigma}$ with $\hat{\sigma}(\ell) = \{\sigma_1, \sigma_2\}$ (for a state ℓ) gives two different functions where $\hat{\sigma}_i(\ell) = \sigma_1$ and $\hat{\sigma}_j(\ell) = \sigma_2$; but these two functions behaves similarly.

A cycle of M^P is a finite sequence $C^P = s_0 \hat{\sigma}_0 s_1 \dots s_{d-1} \hat{\sigma}_{d-1} s_d$ of interleaved cells and symbols such that $\delta^P(s_j, \hat{\sigma}_j) = s_{j+1}$ for all $0 \leq j < d$, and $s_0 = s_d$. Note that, in this definition, d is the length of the cycle C^P . We write $s \in C^P$ if s is one of the cells s_j ($0 \leq j < d$) of the finite sequence of the cycle C^P . A simple cycle is a cycle where all cells s_0, \dots, s_{d-1} are different. We are interested in defining some property on cycles of the perfect-information subset construction for a given MDP.

Definition 2 (Recurrent cyclic sets) Let $C^P = s_0 \hat{\sigma}_0 \dots s_{d-1} \hat{\sigma}_{d-1} s_d$ be a cycle of the perfect-information subset construction M^P for a given MDP M . A recurrent cyclic set for the cycle C^P is a sequence $G = g_0 g_1 \dots g_d$ such that $g_0 = g_d$, and $\emptyset \neq g_i \subseteq s_i$ and $\bigcup_{\ell \in g_i} \text{Post}_{\hat{\sigma}_i(\ell)}(\ell) = g_{i+1}$ for all $0 \leq i < d$.

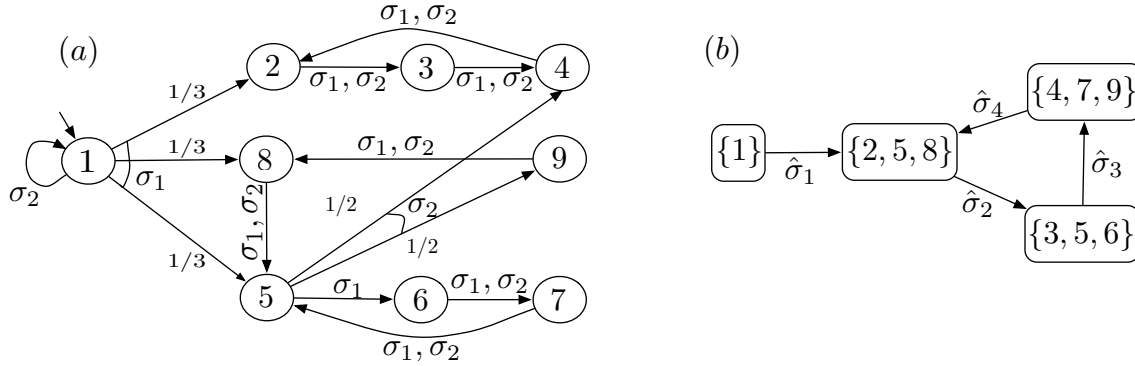


Figure 2: (a) shows an MDP, and (b) shows some part of its perfect information subset construction.

A cycle C^P might have several recurrent cyclic sets. A recurrent cyclic set G for a given cycle C^P , is said to be *minimal* if there is no other recurrent cyclic set G' ($G \neq G'$) such that for $0 \leq i < d$, and for $g_i \in G$, $g'_i \in G'$, we have $g'_i \subseteq g_i$. We denote the set of all minimal recurrent cyclic sets of the cycle C^P by $\Delta(C^P) = \{G \mid G \text{ is a minimal recurrent cyclic set for the cycle } C^P\}$.

Example Consider the MDP M in Figure 2 (the initial distribution is $\mu_0(1) = 1$ and $\mu_0(i) = 0$ for $i \in \{2, \dots, 9\}$). Figure 2(b) shows one cycle of the perfect information subset construction M^P . Let us present $\hat{\Sigma}$ in the table below. Each row labeled by a function $\hat{\sigma}_i$ ($i \in \{1, \dots, 4\}$), each column labeled by a state ℓ ; and each entry shows the value of $\hat{\sigma}_i(\ell)$.

	1	2	3	4	5	6	7	8	9
$\hat{\sigma}_1$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_2$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	σ_1	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_3$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	σ_2	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$
$\hat{\sigma}_4$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$	$\{\sigma_1, \sigma_2\}$

For the cycle $C^P = \{2, 5, 8\} \xrightarrow{\hat{\sigma}_2} \{3, 5, 6\} \xrightarrow{\hat{\sigma}_3} \{4, 7, 9\} \xrightarrow{\hat{\sigma}_4} \{2, 5, 8\}$, the set of minimal recurrent cyclic sets is $\Delta(C^P) = \{\{\{2\}, \{3\}, \{4\}\}, \{\{5\}, \{6\}, \{7\}\}\}$. The elements of $\Delta(C^P)$ are not comparable.

The blind subset construction for an MDP is a special case of its perfect information subset construction where the action functions $\hat{\sigma} \in \hat{\Sigma}$ are restricted to constant functions. In each cell, all states have to choose the same action.

Definition 3 (Blind subset construction of an MDP) The blind subset construction for a given MDP $M = \langle L, \mu_0, \Sigma, \delta \rangle$ is an automaton $M^B = \langle \mathcal{L}, L_I, \Sigma, \delta^B \rangle$ where $\mathcal{L} = \mathcal{P}(L) \setminus \{\emptyset\}$, $L_I = \text{Supp}(\mu_0)$, and for all $s_1, s_2 \in \mathcal{L}$ and $\sigma \in \Sigma$, we have $\delta^B(s_1, \sigma) = s_2$ where $s_2 = \text{Post}_\sigma(s_1)$.

We denote cycles in the blind subset construction by C^B .

3 Synchronizing Objectives for Perfect-Information Strategies

We have defined a perfect-information one-player stochastic game in which the player can see the current state of the game and record the sequence of visited states. We show that synchronizing strategies can be

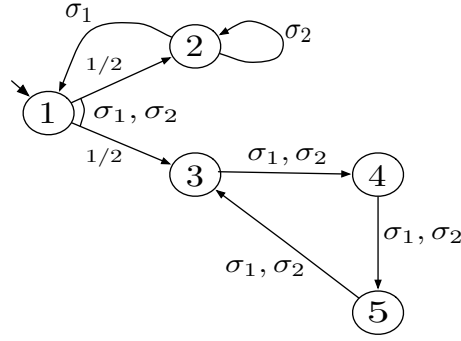


Figure 3: An MDP where memory is necessary to win the strongly synchronizing objective.

characterized in the perfect-information subset construction, giving a decidability result. We also show in the next example that memory may be necessary.

Example Consider the MDP M in Figure 3 (the initial distribution is $\mu_0(1) = 1$ and $\mu_0(i) = 0$ for $i \in \{2, \dots, 5\}$), and let α be the strategy defined as follows: $\alpha((L \times \Sigma)^* \ell)(\sigma) = 1/2$ for all $\sigma \in \Sigma$ and $\ell \in \{1, 3, 4, 5\}$, and for the histories ending in the state 2,

$$\alpha((L \times \Sigma)^* \ell \Sigma 2)(\sigma) = \begin{cases} 1 & \text{if } \ell = 1 \text{ and } \sigma = \sigma_2, \\ 1 & \text{if } \ell \neq 1 \text{ and } \sigma = \sigma_1, \\ 0 & \text{otherwise.} \end{cases}$$

In this example, it is easy to check that the strategy α is strongly synchronizing. In state 2, it plays σ_1 and σ_2 in alternation in order to ensure synchronization with the cycle 3, 4, 5 of length 3. However, none of the memoryless strategies is strongly synchronizing, showing that memory is necessary. This example also shows that memory is necessary for weakly synchronizing objective, as well as for blind strategies.

Proposition 2 *For both strongly and weakly synchronizing objectives, memoryless strategies are not sufficient in MDPs.*

Theorem 1 *For a perfect information game over an MDP M , there exists a strategy α such that M with strategy α is **strongly synchronizing**, if and only if the perfect-information subset construction M^P for M , has an accessible cycle C^P such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$ and for all $g \in G$, $|g| = 1$.*

Proof *Sufficient condition.* We suppose that the perfect-information subset construction M^P for M , has an accessible cycle $C^P = s_0 \hat{\sigma}_0 \dots s_d$ such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$ and for all $g \in G$, we have $|g| = 1$. Since this cycle is accessible, there exists a finite path $P = p_0 \hat{\sigma}'_0 p_1 \dots p_{m-1} \hat{\sigma}'_{m-1} p_m$ in M^P from $p_0 = L_I$ to $p_m = s_0 = s_d$ (See Figure 4). Consider the pure strategy α as follows

$$\alpha((L \times \Sigma)^k \ell) = \begin{cases} \hat{\sigma}'_k(\ell) & \text{if } 0 \leq k < m, \\ \hat{\sigma}'_{(k-m) \bmod d}(\ell) & \text{if } m \leq k. \end{cases}$$

Let us construct a finite Markov chain M' in a way that its long term behavior simulates the long term behavior of the MDP M under the strategy α for synchronizing objectives. This Markov chain is

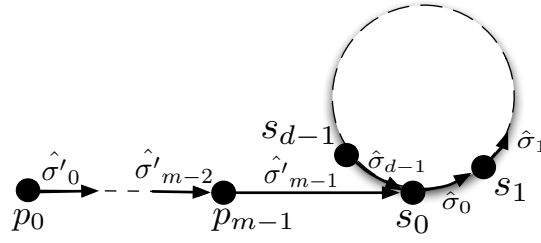


Figure 4: An accessible cycle C^P of M^P which is reachable by a finite path p_0, \dots, p_m .

$M' = (L', \mu'_0, \delta')$ where $L' = \{(i, \ell) \mid 0 \leq i < (m+d) \text{ and } \ell \in L\}$, the initial distribution μ'_0 is defined as follows

$$\mu'_0((i, \ell)) = \begin{cases} \mu_0(\ell) & \text{if } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the probability transition function δ' is defined as follows

$$\delta'((i, \ell))((i', \ell')) = \begin{cases} \delta(\ell, \hat{\sigma}'_i(\ell))(\ell') & \text{if } (0 \leq i < m), (i' = i+1), (\ell \in p_i) \text{ and } (\ell' \in p_{i'}), \\ \delta(\ell, \hat{\sigma}_{i-m}(\ell))(\ell') & \text{if } (m \leq i < m+d), (i' = m + (i - m + 1) \bmod d), \\ & (\ell \in s_{i-m}) \text{ and } (\ell' \in s_{i'-m}), \\ 0 & \text{otherwise.} \end{cases}$$

The idea is that each cell p_i ($0 \leq i < m$) of the path P and, similarly, each cell s_i ($m \leq i < m+d$) of the cycle C^P corresponds to $|L|$ states in the Markov chain M (one for each state of the MDP M). The value of $\delta'((i, \ell))((i', \ell'))$ shows the probability to reach in one step, the state (i', ℓ') from the state (i, ℓ) ; semantically it gives the probability to go from ℓ to ℓ' at step i . We show that (a) if the Markov chain M' is strongly synchronizing, then so is the MDP M under the strategy α and that (b) M' is strongly synchronizing.

Proving (a) is straightforward from the definition of the Markov chain M' . Each state of the MDP M corresponds to $m+d$ state of M' . Then if, from some point, the mass of probability accumulates in one state of M' and afterward moves totally to another one, it happens also in M . In detail, let the sequence X_i^α ($i \in \mathbb{N}$) denote the outcome of the MDP M under the strategy α , and X'_i ($i \in \mathbb{N}$) denote the probability distribution at step i generated by the Markov chain M' . Note that X^α is a random variable over $|L|$ entries, but X' is over $|L| \cdot (m+d)$ entries which has at most $|L|$ non-zero entries. Let us compute and compare the non-zero entries of these two random variable sequences. For $\ell \in L$:

$X_0^\alpha(\ell) = \mu_0(\ell) = X'_0((0, \ell))$ and we have $X'_0((j, \ell)) = 0$ for all $j \neq 0$.

$X_1^\alpha(\ell) = \sum_{\ell' \in L} \mu_0(\ell') \cdot \delta(\ell', \alpha(\ell'))(\ell) = \sum_{\ell' \in L} \mu_0(\ell') \cdot \delta(\ell', \hat{\sigma}'_0(\ell'))(\ell) = \sum_{\ell' \in L} \mu_0(\ell') \cdot \delta'((0, \ell'))((1, \ell)) = X'_1((1, \ell))$ and we have $X'_1((j, \ell)) = 0$ for all $j \neq 1$.

In the next step, let us compute these random variables for $i < m$:

$$\begin{aligned} X_i^\alpha(\ell) &= \\ \sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta(\ell_0, \alpha(\ell_0))(\ell_1) \cdot \delta(\ell_1, \alpha(\ell_0, \alpha(\ell_0), \ell_1))(\ell_2) \cdots \delta(\ell_{i-1}, \alpha(\ell_0, \alpha(\ell_0), \ell_1, \dots, \ell_{i-1}))(\ell) &= \\ \sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta(\ell_0, \hat{\sigma}'_0(\ell_0))(\ell_1) \cdot \delta(\ell_1, \hat{\sigma}'_1(\ell_1))(\ell_2) \cdots \delta(\ell_{i-1}, \hat{\sigma}'_{i-1}(\ell_{i-1}))(\ell) &= \\ \sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta'((0, \ell_0))((1, \ell_1)) \cdot \delta'((1, \ell_1))((2, \ell_2)) \cdots \delta'((i-1, \ell_{i-1}))((i, \ell)) &= X'_i((i, \ell)). \end{aligned}$$

We, also, have $X'_i((j, \ell)) = 0$ for all $j \neq i$, these results give $\|X_i^\alpha\| = \|X'_i\|$ for $i < m$. At the end, consider $i \geq m$:

$$\begin{aligned} X_i^\alpha(\ell) &= \sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta(\ell_0, \alpha(\ell_0))(\ell_1) \cdot \delta(\ell_1, \alpha(\ell_0 \alpha(\ell_0) \ell_1))(\ell_2) \cdots \\ &\delta(\ell_{i-1}, \alpha(\ell_0 \alpha(\ell_0) \ell_1 \dots \ell_{i-1}))(\ell) = \sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta(\ell_0, \hat{\sigma}'_0(\ell_0))(\ell_1) \cdot \delta(\ell_1, \hat{\sigma}'_1(\ell_1))(\ell_2) \cdots \\ &\delta(\ell_{m-1}, \hat{\sigma}'_{m-1}(\ell_{m-1}))(\ell_m) \cdot \delta(\ell_m, \hat{\sigma}_0(\ell_m))(\ell_{m+1}) \cdots \delta(\ell_{i-1}, \hat{\sigma}_{(i-m) \bmod d}(\ell_{i-1}))(\ell) = \\ &\sum_{\ell_0, \ell_1, \dots, \ell_{i-1} \in L} \mu_0(\ell_0) \cdot \delta'((0, \ell_0))((1, \ell_1)) \cdot \delta'((1, \ell_1))((2, \ell_2)) \cdots \delta'((m-1, \ell_{m-1}))((m, \ell_m)) \cdots \delta'((m + \\ &(i-m) \bmod d, \ell_{i-1}))((m + (i-m) \bmod d, \ell)) = X'_i((m + (i-m) \bmod d, \ell)). \end{aligned}$$

We, also, have $X'_i((j, \ell)) = 0$ for all $j \neq i$, this results give $\|X_i^\alpha\| = \|X'_i\|$ for $i \geq m$. We have shown that $X_i^\alpha(\ell) = X'_i((j, \ell))$ where for $0 \leq i < m$, we have $j = i$, and for $i \geq m$, we have $j = m + (i-m) \bmod d$. This simply gives $\|X_i^\alpha\| = \|X'_i\|$ for $i \in \mathbb{N}$; meaning that if the Markov chain M' is synchronizing, so is the MDP M under the strategy α .

To show (b), we study transient and recurrent states of the Markov chain M' . Suppose that $G \in \Delta(C^P)$ is the only recurrent cyclic set of the cycle, and it includes d elements as g_0, \dots, g_{d-1} . Let R be the set of states $(m+i, \ell)$ such that $\ell \in g_i$, for $0 \leq i < d$. We claim that the states of R are the only recurrent states in the Markov chain M' .

- First, we can see that the states of R are recurrent. By construction, the states of R are strongly connected. In addition, we have to prove that if $(m+i, \ell) \in R$ and $(m+i, \ell) \rightarrow (m+j, \ell')$, then $(m+j, \ell') \in R$. This holds by induction on the equality $\cup_{\ell \in g_i} \text{Post}_{\sigma_i(\ell)}(\ell) = g_{i+1}$. Note that $(m+i, \ell) \in R$ implies that $\ell \in g_i$; and if $(m+i, \ell) \rightarrow (m+j, \ell')$ then ℓ' has to lie in g_j .
- Now, we show that the states of R are the **only** recurrent states. By contradiction, suppose that there is another set R' of recurrent states in the Markov chain M' . By Proposition 1 and since the states (i, ℓ) ($0 \leq i < m$) are visited only once, then they could not be recurrent; therefore we discuss only on the states $(m+i, \ell)$ with $0 \leq i < d$ of the Markov chain M' . Let g'_i denote all states included in $\{\ell \mid (m+i, \ell) \in R'\} \cap s_i$ for $0 \leq i < d$. The construction of the Markov chain implies that a state $(m+i, \ell)$ can only have outgoing edges toward some states $(m+(i+1) \bmod d, \ell')$; hence $g'_i \neq \emptyset$ for all $0 \leq i < d$. On the other hand, the definition of recurrent states requires that each accessible state from $(m+i, \ell) \in R'$ could access $(m+i, \ell)$, therefore $\cup_{\ell \in g'_i} \text{Post}_{\sigma_i(\ell)}(\ell) = g'_{i+1}$. It is a contradiction with $|\Delta(C^P)| = 1$.

Based on Proposition 1, for the transient states (k, ℓ) , the probability $X_n((k, \ell))$ vanishes for $n \rightarrow \infty$. Since for all $g \in G$, we have $|g| = 1$, the support of X_n ($n > m$) contains only one recurrent state. Thus, the probability mass accumulates in that state: for all $\varepsilon > 0$, for all $n > n_0$ there is a state (i, ℓ) with $X_n((i, \ell)) > 1 - \varepsilon$, that is $\|X_n^\alpha\| > 1 - \varepsilon$. Hence, $\lim_{n \rightarrow \infty} \|X_n^\alpha\| = 1$ and M' is strongly synchronizing. Therefore, so is the MDP M under the strategy α .

Necessary condition. Assume that the MDP M with strategy α is strongly synchronizing. Then $\forall \varepsilon > 0 \cdot \exists n_0 \in \mathbb{N} \cdot \forall n \geq n_0 \cdot \exists q_n$ such that $X_n^\alpha(q_n) > 1 - \varepsilon$. Moreover the state q_n is unique, and we show below that it is independent of ε (assuming $\varepsilon < \frac{1}{2}$).

Let v be the smallest probability among all probability distributions of the MDP M (i.e., $v = \min_{\ell \in L, \sigma \in \Sigma, \ell' \in \text{Supp}(\delta(\ell, \sigma))} (\delta(\ell, \sigma)(\ell'))$). Let $\varepsilon < \frac{v}{1+v}$. We claim that for all $n \geq n_0$, there exists some action $\sigma \in \Sigma$ such that $\text{Post}_\sigma(q_n) = \{q_{n+1}\}$ is a singleton. Toward contradiction, assume that for all $\sigma \in \Sigma$, the statement $\text{Post}_\sigma(q_n) \neq \{q_{n+1}\}$ is satisfied. The probability which does not inject to q_{n+1} (from q_n) is at least $v \cdot (1 - \varepsilon)$. And since M is strongly synchronizing, we have:

$$1 - \varepsilon \leq \|X_{n+1}^\alpha\| \leq 1 - v \cdot (1 - \varepsilon)$$

This gives $\varepsilon \geq \frac{\nu}{1+\nu}$ which is a contradiction. Therefore, for all $n \geq n_0$, there exists $\sigma \in \Sigma$ such that $Post_\sigma(q_n) = \{q_{n+1}\}$. This implies that the infinite sequence of states $I = q_{n_0} q_{n_0+1} \dots$ is uniquely defined.

The sequence I is used to define a pure synchronizing strategy β from the randomized synchronizing strategy α . This construction implies that the pure strategies are sufficient for strongly synchronizing objectives. We define the pure strategy β as follows:

- for $h \in \text{Hists}(M)$ with $|h| = i$ and $\text{Last}(h) = q_i$, we define $\beta(h) = \sigma$ where $Post_\sigma(\text{Last}(h)) = \{q_{i+1}\}$,
- for $h \in \text{Hists}(M)$ with $|h| = i$ and $\text{Last}(h) \neq q_i$, we define $\beta(h) = \text{Action}(h', i)$
 where $h' \in \text{Hists}(M)$ is the shortest possible history such that (1) $\text{State}(h', i) = \text{Last}(h)$, (2) $Pr^\alpha(h') > 0$, and (3) $\text{Last}(h') = q_j$ with $|h'| = j$. One might notice that a reachable state $\text{Last}(h)$ with a strictly positive probability $Pr^\alpha(h) > 0$, has to access a state of I (such as $\text{Last}(h') = q_j$ where $|h'| = j$); otherwise the MDP M with strategy α , would not be strongly synchronizing. Consequently, the history h' defined above always exists.

As a result, we can define $\text{SizePath}(h) = |h'| - |h|$ to be the size of shortest path from $\text{Last}(h)$ to the infinite sequence I . Note that for h with $|h| = i$ and $\text{Last}(h) = q_i$, we define $\text{SizePath}(h) = 1$. It is easy to see that the MDP M with pure strategy β is also strongly synchronizing.

In the following, we show that there exists a cycle C^P of M^P which has only one recurrent cyclic set G , and all $g \in G$ are singleton. By construction, we have $\beta(h) = \beta(h')$ for all histories $h, h' \in \text{Hists}(M)$ with $\text{Last}(h) = \text{Last}(h')$ and $|h| = |h'|$. Therefore the pure strategy β induces an infinite path P_β , in the perfect-information subset construction M^P . Since the state space of M^P is finite, some cell S has to be visited infinitely many times along P_β . The path between two visits to S along P_β is a cycle (not necessarily a simple cycle) of M^P . We study one of these cycles (starting at S and coming back there), and prove that this cycle satisfies the conditions of the theorem.

Let $\text{Inf}(I)$ denote the set of all states visited infinitely often along I . Hence, there exists $N_{\text{Inf}} \geq n_0$ such that $\forall i \geq N_{\text{Inf}} : q_i \in I \Rightarrow q_i \in \text{Inf}(I)$. Let K_1 be the first step after N_{Inf} in which the path P_β visits S . Let $\text{MaxPath} = \max_{h \in \text{Hists}(M), Pr^\beta(h) > 0, |h| = K_1} (\text{SizePath}(h))$, be the length of the longest path (among the shortest ones) from one reachable state at step K_1 , to the infinite sequence I .

Let C^P be the cycle starting in S at step K_1 , and coming back to this state in some step $K_2 > K_1 + \text{MaxPath}$. We claim that this cycle C^P has only one recurrent cyclic set G , and all subsets $g \in G$ are singleton:

1. $G = \{\{q_i\} \mid q_i \in I \text{ for } K_1 \leq i \leq K_2\}$ is a recurrent cyclic set. We already have proved that there exists $\sigma \in \Sigma$ such that $Post_\sigma(q_n) = \{q_{n+1}\}$ ($n \geq n_0$). Note that for state q_n , the action σ is chosen by the cycle.
2. G is the only recurrent cyclic set. Each state included in S reaches, in at most MaxPath steps, one state of I . Hence, the cell S , as the first element of C^P , cannot have another subset g' constructing another recurrent cyclic set.

We have proved that for a strongly synchronizing MDP M , the perfect information subset construction for M , has a cycle C^P such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$ and for all $g \in G$, $|g| = 1$.

□

Through the proof of Theorem 1, we have seen that for all strategies α such that an MDP M with the strategy α is strongly synchronizing, there is a pure strategy that satisfies the strongly synchronization condition. We will see that this is also the case for weakly synchronizing objective (see the proof of Theorem 2).

Corollary 1 *For both strongly and weakly synchronizing objectives, pure strategies are sufficient in MDPs.*

Theorem 2 *For a perfect information game over an MDP M , there exists a strategy α such that M with strategy α is **weakly synchronizing**, if and only if the perfect-information subset construction M^P for M , has an accessible cycle C^P such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$, there exists $g \in G$ such that $|g| = 1$.*

Proof *Sufficient condition.* We suppose that the perfect-information subset construction M^P for M , has an accessible cycle C^P such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$, there exists $g \in G$ such that $|g| = 1$. Consider a pure strategy similarly to which presented in proof of Theorem 1. Let us, here as well, construct the Markov chain M' , and therefore discuss on transient and recurrent states of M' .

Suppose that $G \in \Delta(C^P)$ is the only recurrent cyclic set of the cycle, and it includes d elements as g_0, \dots, g_{d-1} . Let R be the set of states $(m+i, \ell)$ such that $\ell \in g_i$, for $0 \leq i < d$. As we have shown in proof of Theorem 1, the states of R are the only recurrent states in the Markov chain M' . Let p_n be the probability to be in one state of R at step n . Based on Proposition 1, for the transient states (i, ℓ) the probability $X_n((i, \ell))$ vanishes for $n \rightarrow \infty$, which leads to $\lim_{n \rightarrow \infty} p_n = 1$. On the other hand, by hypothesis, for $G \in \Delta(C^P)$ there exists $g_j \in G$ ($0 \leq j < d$) such that $|g_j| = 1$. Then every d steps, at least once, the probability $p_{m+k \cdot d+j}$ gathers in only one state $(m+j, \ell)$ where $\ell \in g_j$. As a result, for all $k \in \mathbb{N}$, $\max(\|X_{m+d \cdot k}^\alpha\|, \|X_{m+d \cdot k+1}^\alpha\|, \dots, \|X_{m+d \cdot k+d-1}^\alpha\|) \geq p_{m+k \cdot d+j}$. We have shown that $\lim_{n \rightarrow \infty} p_n = 1$, hence $\limsup_{n \rightarrow \infty} \|X_n^\alpha\| = 1$.

Necessary condition. Assume that the MDP M with strategy α is weakly synchronizing meaning that $\limsup_{n \rightarrow \infty} \|X_n^\alpha\| = 1$. Therefore there exists a subsequence $\|X_{i_k}^\alpha\|$ of $\|X_i^\alpha\|$ which approaches to 1 (i.e., $\lim_{k \rightarrow \infty} \|X_{i_k}^\alpha\| = 1$), where $i_0 < i_1 < i_2 < \dots$ is an increasing sequence of indices. Then, for $\varepsilon < 1/2$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ there exists a (unique) state ℓ such that $X_{i_n}^\alpha(\ell) > 1/2$. Let (ℓ, i_n) refer to this unique state at position i_n . Let Inf be the set of all states ℓ such that $X_{i_n}^\alpha((\ell, i_n)) > 1/2$ for infinitely many $n \in \mathbb{N}$.

Hence, there exists $N_{\text{Inf}} \geq n_0$ such that $\forall n \geq N_{\text{Inf}} : X_{i_n}^\alpha((\ell, i_n)) > 1/2 \Rightarrow \ell \in \text{Inf}$.

Since the state space of the MDP is finite, for a specific $q \in \text{Inf}$, we can define a subsequence $(j_k)_{k \in \mathbb{N}}$ of $(i_k)_{k \in \mathbb{N}}$ such that

1. $j_0 \geq N_{\text{Inf}}$, and
2. $X_{j_k}^\alpha((q, j_k)) > 1/2$, and
3. $\text{Supp}(X_{j_k}^\alpha) = \text{Supp}(X_{j_{k+1}}^\alpha)$; in the sequel, we denote to this set by S .

Let (q, j_k) refer to the state q at specific step j_k , and J be the sequence of this states. Note that since j_k is a subsequence of i_k , we have $\lim_{k \rightarrow \infty} \|X_{j_k}^\alpha\| = 1$ as well.

We use the infinite sequence J to construct a winning pure strategy from the winning randomized strategy α . Consider the pure strategy β as follows. For $h \in \text{Hists}(M)$ with $|h| = i$, we define $\beta(h) = \text{Action}(h', i)$ where

$h' \in \text{Hists}(M)$ is the shortest possible history such that (1) $Pr^\alpha(h') > 0$, (2) $\text{Last}(h') = (q, j_k)$ where $|h'| = j_k$ for some $k \in \mathbb{N}$, and in addition (3) $\text{State}(h', i) = \text{Last}(h)$. One might notice that a reachable state $\text{Last}(h)$ with a strictly positive probability $Pr^\alpha(h) > 0$, has to access the infinite sequence J ; otherwise the MDP M with strategy α would not be weakly synchronizing. Consequently, the history h' defined above always exists.

Similarly to the case of strongly synchronizing, we can define $\text{SizePath}(h) = |h'| - |h|$ to be the size of shortest path from $\text{Last}(h)$ to the infinite sequence J .

In the following, we show that for a weakly synchronizing MDP M , there exists a cycle C^P of M^P which has only one recurrent cyclic set G , and there exists $g \in G$ which is singleton. By construction, we have $\beta(h) = \beta(h')$ for all histories $h, h' \in \text{Hists}(M)$ with $\text{Last}(h) = \text{Last}(h')$ and $|h| = |h'|$. Therefore the pure strategy β induces an infinite path P_β in the perfect-information subset construction M^P . The construction of β , also implies that the cell S is visited infinitely many times along P_β . The path taken between two visits to S along P_β is a cycle (not necessarily a simple cycle) of M^P . We study one of these cycles (starting at S and coming back there), and prove that this cycle satisfies the conditions of the theorem.

Let K_1 to be the first step after N_{Inf} in which the path P_β visits S . Let us define $\text{MaxPath} = \max_{h \in \text{Hists}(M), \text{Pr}(h) > 0, |h| = K_1} (\text{SizePath}(h))$ to be the length of the longest path (among the shortest ones) from a reachable state at step K_1 to the infinite sequence J .

Let C^P be the cycle starting in S at step K_1 , and coming back to this state in some step $K_2 > K_1 + \text{MaxPath}$. For convenience, let $d = K_2 - K_1$ denote the length of the cycle C^P . We define the winning pure strategy β' from the strategy β as follows.

- for $h \in \text{Hists}(M)$ with $|h| < K_1 + K_2$, we define $\beta'(h) = \beta(h)$.
- for $h \in \text{Hists}(M)$ with $|h| > K_1 + K_2$, we define $\beta'(h) = \beta(h')$ where $|h| = d \cdot m + |h'|$ for some $m \in \mathbb{N}$, and h' is a history with $K_1 \leq |h'| \leq K_1 + K_2$ and $\text{Last}(h) = \text{Last}(h')$.

In fact, the path corresponding to the strategy β' first reaches the cycle C^P , and then forever follows this cycle. The strategy β' as well as the strategy β is weakly synchronizing. We claim that this cycle $C^P = s_0 \hat{\sigma}_0 \cdots s_d (s_0 = s_d)$ has only one recurrent cyclic set G , and there exists $g \in G$ which is singleton:

1. First we prove that this cycle has one recurrent cyclic set. The size of the cycle is more than MaxPath which shows that some elements of the infinite sequence J are visited along the cycle. Suppose that $(q, j_{K'})$ is the last visited state of J along the cycle, and $K' = j_{K'} - K_1$ is the index of cell $s_{K'}$ including this state. Let us construct a singleton subset $g_{K'} = \{(q, K')\}$. By induction, let $g_{(K'+i+1) \bmod d} = \bigcup_{\ell \in g_{(K'+i) \bmod d}} \text{Post}_{\hat{\sigma}_{(K'+i) \bmod d}}(\ell)$ for all $0 \leq i < d$. Note that, for $i = d - 1$, the set $g_{K'}$ is computed. By definition, the set $G = \{g_0, g_1, \dots, g_{d-1}\}$ is a recurrent cyclic set, if after the computation, we still have $g_{K'} = \{(q, K')\}$.

We claim that $g_{K'} = \{(q, K')\}$. By contradiction, suppose that $g_{K'} \neq \{(q, K')\}$ is satisfied. We have $\limsup_{n \rightarrow \infty} \|X_n^{\beta'}\| = 1$. Then $\forall \varepsilon > 0 \cdot \exists n_0 \in \mathbb{N} \cdot \forall n \geq n_0 \cdot \exists \ell$ such that $X_n^{\beta'}(\ell) > 1 - \varepsilon$. On the other hand, by definition of J , we know that the mass of probability in states of J are more than $1/2$, and in addition we know that all states of the cycle inject probability to J ; these show that the visited states of J along the cycle are candidates to concentrate the probability mass. Let v be the smallest probability among all probability distributions of the MDP M (i.e., $v = \min_{\ell \in L, \sigma \in \Sigma, \ell' \in \text{Supp}(\delta(\ell, \sigma))} (\delta(\ell, \sigma)(\ell'))$). Let $\varepsilon < \frac{v^d}{1+v^d}$, and $X_n^{\beta'}((q, K')) > 1 - \varepsilon$ where $n > n_0$. The probability which does not inject to (q, K') (from (q, K') after d steps), is at least $v^d \cdot (1 - \varepsilon)$. We have:

$$1 - \varepsilon \leq X_{n+d}^{\beta'}((q, K')) \leq 1 - v^d \cdot (1 - \varepsilon)$$

This gives $\varepsilon \geq \frac{v^d}{1+v^d}$ which is a contradiction. Therefore, we have constructed a recurrent cyclic set G for the cycle, and have shown that one element of G is singleton.

2. We can see that G is the only recurrent cyclic set. Each state included in S reaches, at most in MaxPath steps, one state of J . Hence, the cell S , as the first element of C^P , can not have another subset g' constructing another recurrent cyclic set.

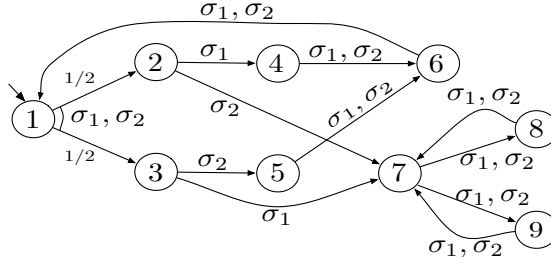


Figure 5: A weakly synchronizing MDP.

We have proved that for a weakly synchronizing MDP M , the perfect information subset construction for M , has a cycle C^P such that $|\Delta(C^P)| = 1$, and for $G \in \Delta(C^P)$, there exists $g \in G$ such that $|g| = 1$.

□

Example The MDP depicted in Figure 5 (the initial distribution is $\mu_0(1) = 1$ and $\mu_0(i) = 0$ for $i \in \{2, \dots, 9\}$) with strategy α which defined as follows $\alpha((L \times \Sigma)^* L)(\sigma) = 1/2$ for $\sigma \in \Sigma$, is weakly synchronizing. Note that $L = \{1, \dots, 9\}$, $\Sigma = \{\sigma_1, \sigma_2\}$.

4 Synchronizing objectives for Blind strategies

We have defined a blind one-player stochastic game where the player is not allowed to observe the current state of the game. We use a characterization of synchronizing blind strategies to show that the existence of synchronizing blind strategies can be decided. We first present an example where the player is blind and has a strategy to make the game synchronizing.

Example The MDP depicted in Figure 6 (the initial distribution is $\mu_0(1) = 1$ and $\mu_0(i) = 0$ for $i \in \{2, \dots, 9\}$) with blind strategy α which defined as following $\alpha((L \times \Sigma)^* L)(\sigma) = 1/2$ for $\sigma \in \{\Sigma\}$ is strongly synchronizing. Note that $L = \{1, \dots, 8\}$, $\Sigma = \{\sigma_1, \sigma_2\}$.

Theorem 3 For a blind game over an MDP M , there exists a strategy α such that M with strategy α is **strongly synchronizing**, if and only if the blind subset construction M^B for M , has an accessible cycle C^B such that $|\Delta(C^B)| = 1$, and for $G \in \Delta(C^B)$ and for all $g \in G$, $|g| = 1$.

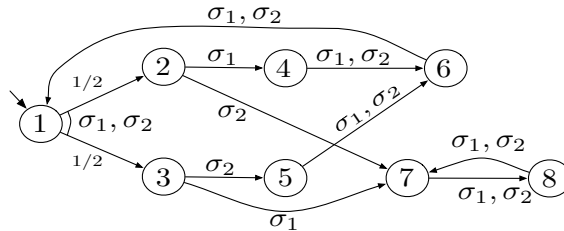


Figure 6: An MDP that with some blind strategy is strongly synchronizing.

Proof Sufficient condition. We suppose that the blind subset construction $M^B = \langle \mathcal{L}, L_I, \Sigma, \delta^B \rangle$ for M , has an accessible cycle C^B such that $|\Delta(C^B)| = 1$, and for $G \in \Delta(C^B)$ and for all $g \in G$, $|g| = 1$. Since this cycle is accessible, then there exists a finite path $P = p_0 \sigma'_0 \dots \sigma'_{m-2} p_{m-1} \sigma'_{m-1} p_m$ in M^B from $p_0 = L_I$ to $p_m = s_0$. Consider the pure blind strategy α as follows

$$\alpha(k) = \begin{cases} \sigma'_k & \text{if } 0 \leq k < m, \\ \sigma'_{(k-m) \bmod d} & \text{if } m \leq k. \end{cases}$$

Let us, construct a Markov chain M' similar to which presented in proof of Theorem 1, with the below probability function: the probability transition function δ' is defined as follows

$$\delta'((i, \ell))((i', \ell')) = \begin{cases} \delta(\ell, \sigma'_i)(\ell') & \text{if } (0 \leq i < m), (i' = i + 1), (\ell \in p_i) \text{ and } (\ell' \in p_{i'}), \\ \delta(\ell, \sigma'_{i-m})(\ell') & \text{if } (m \leq i < m + d), (i' = m + (i - m + 1) \bmod d), \\ & (\ell \in s_{i-m}) \text{ and } (\ell' \in s_{i'-m}), \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that $G \in \Delta(C^B)$ is the only recurrent cyclic set of the cycle, and it includes d elements as g_0, \dots, g_{d-1} . Let R be the set of states $(m + i, \ell)$ such that $\ell \in g_i$, for $0 \leq i < d$. As we have shown in proof of Theorem 1, the states of R are the only recurrent states in the Markov chain M' .

Based on Proposition 1, for the transient states (k, ℓ) , the probability $X_n((k, \ell))$ vanishes for $n \rightarrow \infty$. Since for all $g \in G$, we have $|g| = 1$, the support of X_n ($n > m$) contains only one recurrent state. Thus, the probability mass accumulates in that state: for all $\varepsilon > 0$, for all $n > n_0$ there is a state (i, ℓ) with $X_n((i, \ell)) > 1 - \varepsilon$, that is $\|X_n^\alpha\| > 1 - \varepsilon$. Hence, $\lim_{n \rightarrow \infty} \|X_n^\alpha\| = 1$ and M' is strongly synchronizing. Therefore, so is the MDP M under the blind strategy α .

Necessary condition. We benefit from arguments presented in Proof of Theorem 1; but here since the winning strategy is blind, we use blind subset constructions. □

Theorem 4 *For a blind game over an MDP M , there exists a strategy α such that M with strategy α is **weakly synchronizing**, if and only if the blind subset construction M^B for M , has an accessible cycle C^B such that $|\Delta(C^B)| = 1$, and for $G \in \Delta(C^B)$, there exists $g \in G$ such that $|g| = 1$.*

Proof Sufficient condition. We suppose that the blind subset construction $M^B = \langle \mathcal{L}, L_I, \Sigma, \delta^B \rangle$ for M , has an accessible cycle C^B such that $|\Delta(C^B)| = 1$, and for $G \in \Delta(C^B)$, there exists $g \in G$ such that $|g| = 1$.

Consider a pure strategy similarly to which presented in proof of Theorem 1. Let us, here as well, construct the Markov chain M' , and therefore discuss on transient and recurrent states of M' .

Suppose that $G \in \Delta(C^B)$ is the only recurrent cyclic set of the cycle, and it includes d elements as g_0, \dots, g_{d-1} . Let R be the set of states $(m + i, \ell)$ such that $\ell \in g_i$, for $0 \leq i < d$. As we have shown in proof of Theorem 1, the states of R are the only recurrent states in the Markov chain M' . Suppose that p is the probability to be in one state of R at step n . Based on Proposition 1, for the transient states (i, ℓ) , the probability $X_n((i, \ell))$ vanishes for $n \rightarrow \infty$, which leads $\lim_{n \rightarrow \infty} p = 1$. On the other hand, by hypothesis, for $G \in \Delta(C^B)$, there exists $g_j \in G$ ($0 \leq j < d$) such that $|g_j| = 1$. Then every d steps, at least once, the whole of probability p gathers in only one state $(m + j, \ell)$ where $\ell \in g_j$. As a result, for all $k \in \mathbb{N}$, $\max(\|X_{m+d.k}^\alpha\|, \|X_{m+d.k+1}^\alpha\|, \dots, \|X_{m+d.k+d-1}^\alpha\|) > p$. We have shown that $\lim_{n \rightarrow \infty} p = 1$, hence $\limsup_{n \rightarrow \infty} \|X_n^\alpha\| = 1$.

Necessary condition. We benefit from arguments presented in Proof of Theorem 2; but here since the winning strategy is blind, we use blind subset constructions.

□

From the four previous theorems, we obtain the following result.

Theorem 5 *The problem of deciding the existence of a {perfect-information, blind} strategy in MDPs for a {strongly, weakly} synchronizing objective is decidable.*

We have defined a new class of objectives for Markov decision processes, and we have given a decidable characterization of winning strategies for these objectives. Further investigations will be devoted to studying the precise complexity of the problem, establishing memory bounds, and extending this framework to partially-observable MDPs and stochastic two-player games.

References

- [1] A. Aziz, V. Singhal & F. Balarin (1995): *It Usually Works: The Temporal Logic of Stochastic Systems*. In: *Proc. of CAV: Computer Aided Verification*. LNCS 939, Springer, pp. 155–165. Available at http://dx.doi.org/10.1007/3-540-60045-0_48.
- [2] D. Beauquier, A. M. Rabinovich & A. Slissenko (2002): *A Logic of Probability with Decidable Model-Checking*. In: *Proc. of CSL: Computer Science Logic*. LNCS 2471, Springer, pp. 371–402. Available at http://dx.doi.org/10.1007/3-540-45793-3_21.
- [3] Y. Benenson, R. Adar, T. Paz-Elizur, Z. Livneh & e. Shapiro (2003): *DNA molecule provides a computing machine with both data and fuel*. *Proc. National Acad. Sci. USA* 100, pp. 2191–2196. Available at <http://dx.doi.org/10.1073/pnas.0535624100>.
- [4] A. Bianco & L. de Alfaro (1995): *Model Checking of Probabilistic and Nondeterministic Systems*. In: *Proc. of FSTTCS: Foundations of Software Technology and Theoretical Computer Science*. LNCS 1026, Springer-Verlag, pp. 499–513. Available at http://dx.doi.org/10.1007/3-540-60692-0_70.
- [5] J. Filar & K. Vrieze (1997): *Competitive Markov Decision Processes*. Springer-Verlag. Available at <http://www.springer.com/engineering/mathematical/book/978-0-387-94805-8>.
- [6] V. A. Korthikanti, M. Viswanathan, Y. Kwon & G. Agha (2009): *Reasoning about MDPs as transformers of probability distributions*. In: *Proc. of QEST: Quantitative Evaluation of Systems*. IEEE Computer Society, pp. 199–208. Available at <http://dx.doi.org/10.1109/QEST.2010.35>.
- [7] R. Segala (1995): *Modeling and Verification of Randomized Distributed Real-Time Systems*. Ph.D. thesis, MIT. Available at <http://profs.sci.univr.it/~segala/www/phd.html>. Technical Report MIT/LCS/TR-676.
- [8] M. V. Volkov (2008): *Synchronizing Automata and the Cerny Conjecture*. In: *Proc. of LATA: Language and Automata Theory and Applications*. LNCS 5196, Springer, pp. 11–27. Available at http://dx.doi.org/10.1007/978-3-540-88282-4_4.